电大理工
2008 年 6 月　Study of Science and Engineering at RTVU.　第 2 期　总第 235 期

多因素方差分析中数学模型的建立与检验方法

邢　航

阜新高等专科学校（阜新 123000）

摘　要　在实际问题和科学实验中很多现象的发生或变化是多个因素共同作用的结果,因素之间有些相互没有影响，有些是交互影响的，针对不同情况可以通过试验利用数学模型分析和解决这些问题。文章论证了多因素方差分析建模的条件和基本原理以及受各因素影响的显著性检验方法。

关键词　数学模型　方差分析　离差平方和　随机误差

0　引言

方差分析是在随机干扰存在的情况下，把因素变化所产生的影响分离出来进而做出因素变化对研究对象是否有显著性影响的推断。在实际问题中很多现象的变化是多因素共同作用的结果，多因素方差分析是利用数学模型的可分解性，从总变异中分解出条件误差（组间）和随机误差（组内），并进行对比，从中找出影响试验结果的主要因素。我们主要研究二因素方差分析数学模型的建立与显著性检验。二因素方差分析中，分因素之间无交互影响和有交互影响，即在试验中分不重复试验和重复实验两个方面。下面论述无交互影响不重复试验的二因素方差分析数学模型建立及假设检验。

1　建立数学模型

首先假设所有试验数据都来自同一正态总体。

对试验 A、B 两个因素进行考察，二者试验地位平等。A 因素有 a 个不同水平 $A_1, A_2, \cdots, A_a$；B 因素有 b 个不同水平 $B_1, B_2, \cdots, B_b$。A、B 之间无交互作用，对水平的每种组合$(A_i\ B_j)$进行一次独立试验，共得 ab 个试验结果 $X_{ij}(i=1,2,\cdots,a, j=1,2,\cdots,b)$，试验结果所得数据如表 1。

表 1　方差分析样本数据

因素	B B_1	$B_2\cdots$	$B_j\ \cdots$	B_b	平均值 $\bar{X}_{i.}$
A_1	X_{11}	$X_{12}\cdots$	$X_{1j}\cdots$	X_{1b}	$\bar{X}_{1.}$
A_2	X_{21}	$X_{22}\cdots$	$X_{2j}\cdots$	X_{2b}	$\bar{X}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_i	X_{i1}	$X_{i2}\cdots$	$X_{ib}\cdots$	X_{ib}	$\bar{X}_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_a	$X_{a1}\cdots$	$X_{aj}\cdots$	X_{ab}	X_{ab}	$\bar{X}_{a.}$
平均值	$\bar{X}_{\cdot j}$	$\bar{X}_{\cdot 1}\cdots$	$\bar{X}_{\cdot j}\cdots$	$\bar{X}_{\cdot b}$	$\bar{X}=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}X_{ij}$

其中　$X_{\cdot j}=\frac{1}{a}\sum_{i=1}^{a}X_{ij}\ j=1,2,\cdots,b$

$$X_{i\cdot}=\frac{1}{b}\sum_{j=1}^{b}X_{ij}\quad i=1,2,\cdots,a \tag{1}$$

设 X_{ij} 是服从正态分布 $X_{ij}\sim N(\mu_{ij},\sigma^2)$的总体中抽取的样本，假定 A，B 不存在交互作用。

假定　$X_{ij}=\mu_{ij}+\varepsilon_{ij}$　（2）

$\varepsilon_{ij}\sim N(0,\sigma^2)\ (i=1,2\cdots a;\ j=1,2\cdots b)$，

其中 μ_{ij} 表示 A_iB_j 条件下的理论期望值, ε_{ij}

基金项目： 辽宁省教育厅高等学校科学研究基金资助项目（2008D028）

表示随机误差，且相互独立。由(1)得

$$\mu=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} \qquad (3)$$

$$\mu_{\cdot j}=\frac{1}{a}\sum_{i=1}^{a}\mu_{ij}(j=1,2,\cdots,b) \quad \mu_{i\cdot}=\frac{1}{b}\sum_{j=1}^{b}\mu_{ij}(i=1,2,\cdots,a)$$

令　　$\alpha_i=\mu_{i}.-\mu$　，　$\beta_j=\mu._j-\mu$

称 α_i 为因素 A_i 的第 i 个水平的效应，β_j 为因素 β_j 的第 j 个水平的效应，分别表示因素 A、B 的各个水平的影响的程度。显然有关系式

$$\sum_{i=1}^{a}\alpha_i=\sum_{j=1}^{b}\beta_j=0 \qquad （4）$$

将 μ_{ij} 进行分解

$$\mu_{ij}=\mu+\alpha_i+\beta_j+(\mu-\mu_i-\mu_j+\mu)$$

令　　$\delta_{ij}=\mu_{ij}-\mu_i-\mu_j+\mu$

称为 Ai 和 Bj 的交互效应。而对二因素无重复试验方差分析，假设任意 Ai 和 B_j 之间不存在交互效应，即全部 $\delta_{ij}=0$。这样 μij 分解式可写为 $\mu_{ij}=\mu+\alpha_i+\beta_j$

综上所述，可得二因素无重复试验方差的数学模型

$$\begin{cases} X_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij},\ i=1,2,\cdots,a;j=1,2,\cdots,b, \\ \sum_{i=1}^{a}\alpha_i=0,\sum_{j=1}^{b}\beta_j=0 \\ \varepsilon_{ij}\sim N(0,\sigma^2)\text{且相互独立} \end{cases} \qquad （5）$$

其中 μ，σ^2，α，β（$j=1,2,\cdots,a,j=1,2,\cdots,b$），均为未知参数。

2　显著性检验

对于二因素无交互方差数学模型(1.5)的检验主要是检验两个因素 A 与 B 的影响是否显著。

要判断因素 A 的影响是否显著等价于检验假设

$$H_{01}:\alpha_1=\alpha_2=\cdots=\alpha_a=0$$

要判断因素 B 的影响是否显著等价于检验假设

$$H_{02}:\ \beta_1=\beta_2=\cdots=\beta_b=0$$

检验上述假设的基本原理是将总离差平方和分解为各因素导致的离差平方和及随机误差导致的离差平方和。具体方法如下

设定

$$\bar{X}=\frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}X_{ij}$$

$$\bar{X}_{i}\cdot=\frac{1}{b}\sum_{j=1}^{b}X_{ij}(i=1,2,\cdots,a) \qquad （6）$$

$$X._J=\frac{1}{a}\sum_{i=1}^{a}X_{ij}(j=1,2,\cdots,b)$$

由（6）有

$$SS_A=b\sum_{i=1}^{a}(\bar{X}_{i}\cdot-\bar{X})^2$$

$$SS_B=a\sum_{j=1}^{b}(\bar{X}\cdot_j-\bar{X})^2 \qquad (7)$$

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij}-\bar{X}_{i}\cdot-\bar{X}\cdot_j+\bar{X})^2$$

其中 SS_A 称为因素 A 的效应平方和,表示因素 A 的水平变化引起的影响；SS_B 称为因素 B 的效应平方和，表示因素 B 的水平变化引起的影响；SS_B 称为误差平方和，表示试验的随机误差影响。总离差分解后的公式为

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{X}_{ij}-\bar{X})^2=SS_A+SS_B+SS_E \ (8)$$

上式表明总离差的平方和分解为二因素的影响（组间）和随机误差影响（组内）的离差平方和。

在(8)成立时，利用关于正态分布平方和分解的 Cochran 定理。可证明 H_{01} 与 H_{02} 分别成立时的 SS_A，SS_B，SS_E 及 M_{SS} 的分布规律。

Cochran 定理：X_1，X_2，…Xn 为 n 个相互独立的服从标准正态分布的随机变量，$Q_i(i=1,2,\cdots,k)$是某些 X_1，X_2，⋯Xn 线性组合的平方和，其自由度分 SS_A 别为 $di(i=1,2,\cdots,k)$。

如果　$Q_1+Q_2\cdots+Q_k \sim x^2(n)$

且 $d_1+d_2+\cdots+d_k=n$　，

则　$Q_i \sim X^2 \quad (i=1,2,\cdots,k)$

并且 Q , Q_2 , …, Q_k 相互独立。

在（8）成立的条件下，利用 Cochran 分解定理，可证明在仅有 H_{01} 成立时，有

$$\frac{SS_E}{\sigma^2} \sim x^2[(a-1)(b-1)]$$

$$\frac{SS_A}{\sigma^2} \sim x^2(a-1) \tag{9}$$

且它们相互独立，从而有统计量

$$F_A=\frac{M_{SS_A}}{M_{SS_B}}=\frac{SS_A/a-1}{SS_B/(a-1)(b-1)} \sim F[a-1,(a-1)(b-1)] \tag{10}$$

所以对给定的显著性水平 α,查 F 分布表,得临界值 $F_a[a\text{-}1,(a\text{-}1)\ (b\text{-}1)]$ ，当 $F_A > F_a$ 时，拒绝 H_{01}，因素 A 影响显著；反之，接受 H_{01}，因素 A 影响不显著。

同理，可得在仅有 H_{02} 成立时因素 B 影响是否显著的检验方法。

综上所述，可得到二因素无交互影响试验方差分析数学模型显著性假设检验的统计分析结果如表 2。

表 2　无交互影响二因素方差分析统计决断

方差来源	平方和 SS	自由度 d	均方 M_{SS}	F 值	显著性
A 影响	SS_A	a-1	M_{ss}	F_A	$F_A > F_a$ 显著
B 影响	SS_B	b-1	M_{ss}	F_B	$F_B > F_a$ 显著
随机误差	SS_E	(a-1)(b-1)	M_{ss}		
总离差	SS_T	ab-1			

表 2 中的各项指标利用表 1 中的样本数据计算，可使用下面的简捷式

$$SS_A=Q_A-P \qquad SS_B=Q_B-P$$

$$SS_E=R-Q_A-Q_B+P \tag{11}$$

$$SS_T=R-P$$

其中

$$Q_A=\frac{1}{b}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}X_{ij}\right)^2$$

$$Q_B=\frac{1}{a}\sum_{j=1}^{b}\left(\sum_{i=1}^{a}X_{ij}\right)^2$$

$$P=\frac{1}{ab}\left(\sum_{i=1}^{a}\sum_{j=1}^{b}X_{ij}\right)^2$$

$$R=\sum_{i=1}^{a}\sum_{j=1}^{b}(X_{ij})^2$$

3　结语

本文论述了二因素无交互作用方差分析数学摸型的建立方法，并论证了利用 Cochran 分解定理对各因素影响的显著性进行假设检验的方法。事实上还有很多生产实际和科学实验方面的问题是二个以上的因素影响且交互作用的，均可利用数学模型进行分析和检验，其原理是相通的本文不在赘述。

参考文献

[1]欧贵兵,刘清国.概率统计及其应用.北京:科学出版社 2007:200-228.

[2]梅国平,袁捷敏,毛小兵等.概率论与数理统计.北京:科学出版社,2007:234-254.

[3]王松桂,陈敏,陈立萍等.线性统计模型.北京:高等教育出版社,1999:138-161.

[4]王孝玲.教育统计学(第二版).上海:华东师范大学出版社,2001:184-192.

[5]戴明强,李卫军,杨鹏飞.数学模型极其应用.北京:科学出版社,2007:1-9.

[6]郭嗣琮,陈刚.不规则介质采场模糊流的数学模型.辽宁工程技术大学学报.2001,20(5):666-668.